ENUMERATING CONTINGENCY
TABLES VIA RANDOM PERMANENTS

ALEXANDER BARVINOK

March 2006

ABSTRACT. Given m positive integers $R = (r_i)$, n positive integers $C = (c_j)$ such that $\sum r_i = \sum c_j = N$, and mn non-negative weights $W = (w_{ij})$, we consider the total weight $T = T(R, C; W)$ of non-negative integer matrices (contingency tables) $D = (d_{ij})$ with the row sums r_i, column sums c_j, and the weight of D equal to $\prod w_{ij}^{d_{ij}}$. We present a randomized algorithm of a polynomial in N complexity which computes a number $T' = T'(R, C; W)$ such that $T' \leq T \leq \alpha(R, C)T'$ where $\alpha(R, C) = \min \left\{ \prod r_i! r_i^{-r_i}, \ \prod c_j! c_j^{-c_j} \right\} N^N / N!$. In many cases, $\ln T'$ provides an asymptotically accurate estimate of $\ln T$. The idea of the algorithm is to express T as the expectation of the permanent of an $N \times N$ random matrix with exponentially distributed entries and approximate the expectation by the integral T' of an efficiently computable log-concave function on $\mathbb{R}^{mn}$. Applications to counting integer flows in graphs are also discussed.

1. INTRODUCTION AND MAIN RESULTS

(1.1) Contingency tables. Let us fix m positive integers $r_1, \dots, r_m$ and n positive integers $c_1, \dots, c_n$ such that

$$r_1 + \dots + r_m = c_1 + \dots + c_n = N.$$

A non-negative $m \times n$ integer matrix $D = (d_{ij})$ with the row sums $r_1, \dots, r_m$ and the column sums $c_1, \dots, c_n$ is called a *contingency table* with the margins $R = (r_1, \dots, r_m)$ and $C = (c_1, \dots, c_n)$. The problem of efficient enumeration of contingency tables with prescribed margins has attracted a lot of attention recently, see [DG95], [D+97], [CD03], [Mo02], [C+05]. The interest in contingency tables

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

is motivated by applications to statistics, combinatorics and representation theory, cf. [DG95] and [DG04].

Let $W = (w_{ij})$ be an $m \times n$ matrix of non-negative weights w_{ij}. In this paper, we consider the quantity

$$T(R, C; W) = \sum_D \prod_{ij} w_{ij}^{d_{ij}},$$

where the sum is taken over all contingency tables $D = (d_{ij})$ with the given margins $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$. Thus if $w_{ij} = 1$ for all i, j, the value of $T(R, C; W)$ is equal to the number of the contingency tables with the given margins. If $w_{ij} \in \{0, 1\}$, the number $T(R, C; W)$ counts contingency tables D for which we have $d_{ij} = 0$ for all i, j with $w_{ij} = 0$ (here we agree that $0^0 = 1$). In this case, $T(R, C; W)$ can be interpreted as the number of integer flows in a bipartite graph, see [B+04] and [C+05]. We note that counting integer flows in a general graph on n vertices can be reduced to counting of integer flows in a bipartite graph on $n + n$ vertices and hence to counting weighted $n \times n$ contingency tables, see Section 1.5.

Geometrically, one can view $T(R, C; W)$ as the generating function over all integer points in the transportation polytope of $m \times n$ non-negative matrices with the row sums r_i and column sums c_j, cf. [BP99].

We note that if $m = n$, $R = (1, \ldots, 1)$, and $C = (1, \ldots, 1)$ then

$$T(R, C; W) = \operatorname{per} W$$

is the *permanent* of the weight matrix W, that is,

$$\operatorname{per} W = \sum_\pi \prod_{i=1}^n w_{i\pi(i)},$$

where the sum is taken over all bijections $\pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$, cf., for example, [Mi78]. A randomized polynomial time approximation algorithm to compute the permanent of a given non-negative matrix was recently obtained by M. Jerrum, A. Sinclair, and E. Vigoda [J+04].

We show that $T(R, C; W)$ can be represented as the expected permanent of an $N \times N$ random matrix with exponentially distributed entries.

Recall that a random variable γ is *standard exponential* if

$$\mathbf{P}(\gamma \geq t) = \begin{cases} e^{-t} & \text{for } t > 0 \\ 1 & \text{for } t \leq 0. \end{cases}$$

Our starting point is the following result.

2

(1.2) Theorem. *Given positive integers* $r_1, \dots, r_m$ *and* $c_1, \dots, c_n$ *such that*

$$r_1 + \dots + r_m = c_1 + \dots + c_n = N$$

and mn *real numbers* w_{ij}, $i = 1, \dots, m$ *and* $j = 1, \dots, n$, *let us construct the* $N \times N$ *random matrix* A *as follows. The matrix* $A = A(\gamma)$ *is a function of the* $m \times n$ *matrix* $\gamma = (\gamma_{ij})$ *of independent standard exponential random variables* γ_{ij}. *We represent the set of rows of* A *as a disjoint union of* m *subsets* $R_1, \dots, R_m$, *where* $|R_i| = r_i$ *for* $i = 1, \dots, m$ *and the set of columns of* A *as a disjoint union of* n *subsets* $C_1, \dots, C_n$, *where* $|C_j| = c_j$. *Thus* A *is split into* mn *blocks* $R_i \times C_j$. *We sample* mn *independent standard exponential random variables* γ_{ij}, $i = 1, \dots, m$ *and* $j = 1, \dots, n$, *and fill the entries of the block* $R_i \times C_j$ *of* $A = A(\gamma)$ *by the copies* $w_{ij}\gamma_{ij}$. *Then the total weight* $T(R, C; W)$ *of the* $m \times n$ *contingency tables with the row sums* $r_1, \dots, r_m$ *and column sums* $c_1, \dots, c_n$, *where the table* $D = (D_{ij})$ *is counted with the weight*

$$w(D) = \prod_{i,j} w_{ij}^{d_{ij}},$$

is equal to

$$\frac{\mathbf{E} \operatorname{per} A}{r_1! \cdots r_m! c_1! \cdots c_n!}.$$

We prove Theorem 1.2 in Section 2.

Although we can compute individual permanents $\operatorname{per} A$ via the algorithm of [J+04], evaluating the expectation is still a difficult problem. However, the expectation of an *approximate permanent* of A can be computed efficiently.

(1.3) An approximation algorithm to compute $T(R, C; W)$**.** We rely heavily on the theory of *matrix scaling* and its applications to approximating the permanent, in particular as described in [Lo71], [Si64], [KK96], [NR99], [L+00], and [GS02], as well as on the Markov chain based algorithms for integrating log-concave densities [AK91], [F+94], [FK99], and [Ve05].

We assume here that the weights w_{ij} are strictly positive, which is not really restrictive since the zero weights can be replaced by sufficiently small positive weights.

Let $A = (a_{ij})$ be an $N \times N$ positive matrix. Then there exist positive numbers $\xi_1, \dots, \xi_N; \eta_1, \dots, \eta_N$ and a positive doubly stochastic (all row and column sums are equal to 1) $N \times N$ matrix $B = B(A)$, $B = (b_{ij})$, such that

$$a_{ij} = b_{ij} \xi_i \eta_j \quad \text{for} \quad i, j = 1, \dots, N.$$

Moreover, the matrix $B = B(A)$ is unique while the numbers $\xi_1, \dots, \xi_N$ and $\eta_1, \dots, \eta_N$ are unique up to a scaling $\xi_i \longmapsto \xi_i \tau$, $\eta_j \longmapsto \eta_j \tau^{-1}$. This allows us to define the function

$$\sigma(A) = \prod_{i=1}^{N} \xi_i \eta_i.$$

3

We use the two crucial facts about σ:

• There is an algorithm, which, given a positive $N \times N$ matrix A and a number $\epsilon > 0$ computes $\sigma(A)$ within relative error ϵ in time polynomial in $\ln \epsilon^{-1}$ and N [L+00]

and

• The function σ is *log-concave*, that is,

$$\ln \sigma \left(\alpha_1 A_1 + \alpha_2 A_2 \right) \geq \alpha_1 \ln \sigma \left(A_1 \right) + \alpha_2 \ln \sigma \left(A_2 \right)$$

for any positive matrices A_1 and A_2 and any non-negative α_1 and α_2 such that $\alpha_1 + \alpha_2 = 1$.

Our algorithm is based on replacing per A in Theorem 1.2 by the (scaled) function $\sigma(A)$. Namely, we define

$$T'(R, C; W) = \frac{N!}{N^N} \frac{\mathbf{E}\, \sigma(A)}{r_1! \cdots r_m! c_1! \cdots c_n!}.$$

Since both $\sigma(A)$ and the exponential density on $\mathbb{R}^{mn}$ are log-concave, and since $\sigma(A)$ is efficiently computable for any positive A, we can apply results of R. Kannan et al. [AK91], [F+94], and [FK99] and of L. Lovász and S. Vempala [Ve05] on efficient integration of log-concave functions to show that there is a randomized fully polynomial time approximation scheme to compute $T'(R, C; W)$.

• We present a randomized algorithm, which, for any $\epsilon > 0$ computes $T'(R, C; W)$ within relative error ϵ in time polynomial in ϵ^{-1} and N (in the unit cost model).

We discuss the details of our algorithm in Sections 3 and 4. Namely, in Section 3 we present the necessary results regarding $\sigma(A)$ while in Section 4 we discuss the integration problem.

Finally, we discuss how well the value of $T'(R, C; W)$ approximates $T(R, C; W)$.

(1.4) Theorem. *For the number $T'(R, C; W)$ computed by the algorithm of Section 1.3, we have*

$$T'(R, C; W) \leq T(R, C; W) \leq \alpha(R, C) T'(R, C; W),$$

where

$$\alpha(R, C) = \frac{N^N}{N!} \min \left\{ \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}}, \quad \prod_{j=1}^{n} \frac{c_j!}{c_j^{c_j}} \right\}.$$

We prove Theorem 1.4 in Section 5.

4

Let us consider the case of $m = n$ and

$$r_1 = \ldots = r_n = c_1 = \ldots = c_n = t.$$

Thus we are enumerating weighted *magic squares*, that is, $n \times n$ square matrices with row and column sums equal to t. Applying Stirling's formula in Theorem 1.4, we achieve the approximation factor of $\alpha(R, C) \leq (nt)^{-1/2}(const \cdot t)^{n/2}$, that is, simply exponential in the size n of the matrix and polynomial in the line sum t. Thus, for any fixed t, the algorithm of Section 1.3 can be considered as an extension of the algorithm of N. Linial, A. Samorodnitsky, and A. Wigderson [L+00] for computing the permanent of a positive matrix within a simply exponential factor. On the other hand, if n is fixed and t grows, the number of magic squares grows as a polynomial in t of degree $(n-1)^2$, see for example, [St97]. Thus, in this case, the algorithm of Section 1.3 allows us to capture the logarithmic order of $T(R, C; W)$.

Apart from the case of $r_i = c_j = 1$ (computation of the permanent), most of the research thus far dealt with the case of $w_{ij} = 1$, that is, with the non-weighted enumeration of contingency tables. M. Dyer, R. Kannan, and J. Mount [D+97] showed that if the margins are not too small, $r_i = \Omega\left(n^2 m\right)$ and $c_j = \Omega\left(m^2 n\right)$, the Monte Carlo based approach allows one to approximate the number of contingency tables within a prescribed relative error $\epsilon > 0$ in time polynomial in m, n, and ϵ^{-1}. In this case, the number of tables is well approximated by the volume of the transportation polytope of $m \times n$ non-negative matrices with the row sums r_i and the column sums c_j. Subsequently, B. Morris [Mo02] improved the bounds to $r_i = \Omega\left(n^{3/2} m \ln m\right)$ and $c_j = \Omega\left(m^{3/2} n \ln n\right)$. The approximation we get is much less precise, but applies to arbitrary weights $W = (w_{ij})$ and seems to be non-trivial even for $w_{ij} = 1$ and moderate values of r_i, c_j. For example, if $m = n$ and $r_i = c_j = n$, we approximate the number of tables within a factor of $(const \cdot n)^{(n-2)/2}$, while the exact number of tables is at least $e^{O(n^2)}$. In other words, in many non-trivial cases we get an asymptotically accurate estimate of $\ln T(R, C; W)$.

Since every log-concave density can be arbitrarily closely approximated by the push-forward (projection) of the Lebesgue measure restricted to some higher dimensional convex body, the algorithm of Section 1.3 can be viewed as a volume approximation algorithm. In contrast to [D+97] and [Mo02], the convex body whose volume we approximate is not polyhedral.

(1.5) Counting integer flows in a graph. Let $G = (V, E)$ be a directed graph with the set V of vertices and the set E of edges. Hence every edge $e \in E$ is incident to the *head* head$(v) \in V$ of e and the *tail* tail$(e) \in V$. We assume that G is connected and that it does not contain loops or multiple edges. Suppose further that each vertex v has an integer number $a(v)$, called the *excess* of v, assigned to it, and that

$$\sum_{v \in V} a(v) = 0.$$

5

A set of non-negative integers $x(e) : e \in E$ is called an *integer feasible flow* in G if for every $v \in V$ the balance condition holds:

$$\sum_{e: \ head(e)=v} x(e) - \sum_{e: \ tail(e)=v} x(e) = a(v).$$

If G does not contain directed cycles $v_1 \to v_2 \to \ldots \to v_k \to v_1$, the number of integer feasible flows is finite, possibly 0. The problem of efficient counting of integer feasible flows in a given graph has attracted some attention recently, cf. [B+04] and [C+05]. A variation of the problem involves introducing capacities of edges (upper bounds on the flows).

One can express the number of integer feasible flows in a graph with $|V| = n$ vertices as the number $T(R, C; W)$ of weighted $n \times n$ contingency tables, where $w_{ij} \in \{0, 1\}$ for all i, j. To this end, let us construct a bipartite graph with $n + n$ vertices as follows. For every vertex $v \in V$, we introduce the left copy v_L and the right copy v_R. The directed edges $u \to v$ of G are represented by the edges $u_L \to v_R$ of the bipartite graph. We also introduce edges $v_L \to v_R$. Finally, let us choose a sufficiently large integer z, for example,

$$z = \sum_{v: \ a(v)>0} a(v)$$

and let us assign the excesses

$$a\left(v_L\right) = z - a(v) \quad \text{and} \quad a\left(v_R\right) = z.$$

With a feasible flow in the original graph G we associate a feasible flow in the constructed bipartite graph by letting the flow on the edge $u_L \to v_R$ equal to the flow on the edge $u \to v$ and assigning the flow $v_L \to v_R$ so as to satisfy the balance conditions. This correspondence is a bijection between the integer feasible flows in G and the bipartite graph. Hence the number of such flows is equal to the number of weighted $n \times n$ contingency tables with the rows and columns indexed by the vertices $v \in V$, the row margins $z - a(v)$, the column margins z and the matrix $W = (w_{ij})$ of weights defined by $w_{ij} = 1$ for $(i, j) \in E$ and $w_{ij} = 0$ for $(i, j) \notin E$.

2. Proof of Theorem 1.2

If γ is a standard exponential random variable then for any integer $d \geq 0$ we have

$$\mathbf{E} \, \gamma^d = \int_0^{+\infty} \tau^d e^{-\tau} \, d\tau = d!.$$

Let us consider the random matrix $A = (a_{pq})$ as defined in Theorem 1.2. We identify both the set of rows of A and the set of columns of A with the set $\{1, \ldots, N\}$.

For every permutation $\pi : \{1, \ldots, N\} \longrightarrow \{1, \ldots, N\}$, let

$$(2.1) \qquad t_\pi = \prod_{k=1}^{N} a_{k\pi(k)}$$

be the corresponding term of $\operatorname{per} A$. Thus

$$(2.2) \qquad \mathbf{E} \operatorname{per} A = \sum_\pi \mathbf{E}\, t_\pi,$$

where the sum is taken over all permutations π. With every permutation π we associate a contingency table $D = D(\pi)$, called the *pattern* of π as follows. We let $D = (d_{ij})$ where d_{ij} is the number of indices $k \in \{1, \ldots, N\}$ such that $k \in R_i$ and $\pi(k) \in C_j$, so the $(k, \pi(k))$th entry of A lies in the block $R_i \times C_j$ of A.

For the corresponding term t_π of the permanent (2.1), we have

$$(2.3) \qquad \mathbf{E}\, t_\pi = \prod_{ij} w_{ij}^{d_{ij}} d_{ij}!,$$

where $D = (d_{ij})$ is the pattern of π.

Now, let us count how many permutations π have a given pattern $D = (d_{ij})$. Let us represent each subset R_i of rows as a disjoint (ordered) union

$$R_i = \bigcup_{j=1}^{n} R_{ij} \quad \text{for} \quad i = 1, \ldots, m$$

of (possibly empty) subsets R_{ij} with $|R_{ij}| = d_{ij}$ and each subset C_j of columns as a disjoint (ordered) union

$$C_j = \bigcup_{i=1}^{m} C_{ij} \quad \text{for} \quad j = 1, \ldots, n$$

of (possibly empty) subsets C_{ij} with $|C_{ij}| = d_{ij}$. This pair of partitions gives rise to exactly $\prod_{ij} d_{ij}!$ permutation π with the pattern D: we choose π in such a way that if $k \in R_{ij}$ then $\pi(k) \in C_{ij}$ and we note that there are precisely $d_{ij}!$ bijections $R_{ij} \longrightarrow C_{ij}$.

On the other hand, the number of partitions $R_i = \bigcup_j R_{ij}$ is

$$\frac{r_i!}{\prod_{j=1}^{n} d_{ij}!}$$

while the number of partitions $C_j = \bigcup_i C_{ij}$ is

$$\frac{c_j!}{\prod_{i=1}^{m} d_{ij}!}.$$

Therefore, the number of permutations with the given pattern $D = (d_{ij})$ is

$$\frac{r_1! \cdots r_m! c_1! \cdots c_n!}{\prod_{ij} d_{ij}!}.$$

The proof now follows by (2.3) and (2.2). $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark. Let us modify the definition of A as follows: instead of filling the $R_i \times C_j$ block by the copies of $w_{ij}\gamma_{ij}$, we fill $R_i \times C_j$ by the copies of just w_{ij}, so A is constructed deterministically. It follows from the proof above that the value of

$$\frac{\operatorname{per} A}{r_1! \cdots r_m! c_1! \cdots c_n!}$$

is equal to the total weight of the contingency tables with the margins $r_1, \ldots, r_m$ and $c_1, \ldots, c_n$ provided the weight of the table $D = (d_{ij})$ is

$$\prod_{ij} \frac{w_{ij}^{d_{ij}}}{d_{ij}!}$$

(the Fisher-Yates statistics, cf. [DG95]).

For another proof of Theorem 1.2 in a particular case of $w_{ij} = 1$, see [Ba05].

3. MATRIX SCALING

Here we summarize the matrix scaling results that we need. All the results in Theorem 3.1 below can be found in the literature

We reproduce the approach of L. Gurvits and A. Samorodnitsky [GS02] adapted to the case of the permanent (paper [GS02] treats a more general and more complicated setting of mixed discriminants), which is, in turn, a modification of D. London's [Lo71] approach.

Also, we restrict ourselves to the case of strictly positive matrices to avoid dealing with certain combinatorial subtleties.

(3.1) Theorem. *For every positive $N \times N$ matrix $A = (a_{ij})$ there exist unique positive N-vectors $x = x(A)$, $y = y(A)$, and an $N \times N$ positive matrix $B = B(A)$*

$$x = (\xi_1, \ldots, \xi_N), \quad y = (\eta_1, \ldots, \eta_N), \quad and \quad B = (b_{ij})$$

so that the following holds
   (1) *We have*
$$a_{ij} = b_{ij}\xi_i\eta_j \quad for \quad i, j = 1, \ldots, N;$$

   (2) *We have*
$$\prod_{j=1}^{N} \eta_j = 1;$$

8

(3) *Matrix B is doubly stochastic, that is,*

$$\sum_{i=1}^{N} b_{ij} = 1 \quad for \quad j = 1, \ldots, N \quad and \quad \sum_{j=1}^{N} b_{ij} = 1 \quad for \quad i = 1, \ldots, N.$$

Let us define

$$\sigma(A) = \prod_{i=1}^{N} \xi_i, \quad where \quad x(A) = (\xi_1, \ldots, \xi_N).$$

Then σ is a log-concave function on the set of positive matrices:

$$\ln \sigma(\alpha_1 A_1 + \alpha_2 A_2) \geq \alpha_1 \ln \sigma(A_1) + \alpha_2 \ln \sigma(A_2)$$

for any two positive $N \times N$ matrices A_1 and A_2 and any two numbers $\alpha_1, \alpha_2 \geq 0$ such that $\alpha_1 + \alpha_2 = 1$.

Proof. Let us consider the hyperplane

$$H = \left\{ (\tau_1, \ldots, \tau_N) : \quad \sum_{i=1}^{N} \tau_i = 0 \right\}$$

in $\mathbb{R}^N$. With a positive matrix $A = (a_{ij})$, we associate the function $f_A : \mathbb{R}^N \longrightarrow \mathbb{R}$,

$$f_A(t) = \sum_{i=1}^{N} \ln \left(\sum_{j=1}^{N} a_{ij} e^{\tau_j} \right), \quad where \quad t = (\tau_1, \ldots, \tau_N).$$

Then the restriction of $f_A(t)$ on H is strictly convex and, moreover, f_A attains its unique minimum $t^* = (\tau_1^*, \ldots, \tau_N^*)$, $t^* = t^*(A)$, on H, see [GS02].

Since f_A is smooth, t^* is also a critical point and the gradient of f_A at the critical point is proportional to vector $(1, \ldots, 1)$, from which we get

(3.1.1)
$$\sum_{i=1}^{N} \left(\frac{a_{ik} e^{\tau_k^*}}{\sum_{j=1}^{N} a_{ij} e^{\tau_j^*}} \right) = \gamma$$

for some constant γ and $k = 1, \ldots, N$.

Let

$$\xi_i = \sum_{j=1}^{N} a_{ij} e^{\tau_j^*} \quad for \quad i = 1, \ldots, N,$$

let

$$\eta_j = e^{-\tau_j^*} \quad for \quad j = 1, \ldots, N,$$

9

and let us define an $N \times N$ matrix $B = (b_{ij})$ by

$$b_{ij} = \frac{a_{ij}}{\xi_i \eta_j} \quad \text{for} \quad i, j = 1, \dots, N.$$

We note that

$$\prod_{j=1}^{N} \eta_j = 1$$

since t^* lies in the hyperplane H with $\tau_1 + \dots + \tau_N = 0$. Then, by (3.1.1), we have

(3.1.2)
$$\sum_{i=1}^{N} b_{ij} = \gamma \quad \text{for} \quad j = 1, \dots, N$$

On the other hand,

(3.1.3)
$$\sum_{j=1}^{N} b_{ij} = 1 \quad \text{for} \quad i = 1, \dots, N.$$

Since B is a square matrix, comparing (3.1.2) and (3.1.3), we infer that $\gamma = 1$ and so we established the existence of $x = (\xi_1, \dots, \xi_N)$ and $y = (\eta_1, \dots, \eta_N)$ and B satisfying (1)–(3).

To show uniqueness, we note that if $x = (\xi_1, \dots, \xi_N)$, $y = (\eta_1, \dots, \eta_N)$, and B satisfy (1)–(3), then we must have

$$\xi_i = \sum_{j=1}^{N} a_{ij} \eta_j^{-1} \quad \text{for} \quad i = 1, \dots, N$$

and hence, necessarily, the point $t = (\tau_1, \dots, \tau_N)$ defined by

$$\tau_j = -\ln \eta_j \quad \text{for} \quad j = 1, \dots, N$$

is a critical point of $f_A(t)$ on H. Since $f_A(t)$ is strictly convex on H, there is a unique critical point $t^* = t^*(A)$.

Thus function $\sigma(A)$ is well-defined. Moreover, we can write

(3.1.4)
$$\ln \sigma(A) = \sum_{i=1}^{N} \ln \xi_i = f_A(t^*) = \min_{t \in H} f_A(t).$$

We observe that for any fixed t, the function $g(A) = f_A(t)$ is concave on the set of positive matrices $A = (a_{ij})$.

10

Hence for any $t \in H$ and any $\alpha_1, \alpha_2 \geq 0$ such that $\alpha_1 + \alpha_2 = 1$, we have

$$f_{\alpha_1 A_1 + \alpha_2 A_2}(t) \geq \alpha_1 f_{A_1}(t) + \alpha_2 f_{A_2}(t) \geq \alpha_1 \ln \sigma(A_1) + \alpha_2 \ln \sigma(A_2).$$

Taking the minimum over $t \in H$, we conclude that

$$\ln \sigma(\alpha_1 A_1 + \alpha_2 A_2) \geq \alpha_1 \ln \sigma(A_1) + \alpha_2 \ln \sigma(A_2),$$

so $\sigma(A)$ is indeed log-concave. $\square$

(3.2) Remark. Another useful property of $\sigma(A)$ which easily follows from (3.1.4) is *monotonicity*: if $A = (a_{ij})$ and $A' = (a'_{ij})$ are positive matrices such that $a'_{ij} \leq a_{ij}$ for all i and j then $\sigma(A') \leq \sigma(A)$. We also note that $\sigma(A)$ is positive homogeneous of degree N: $\sigma(\lambda A) = \lambda^N \sigma(A)$ for all positive $N \times N$ matrices A and all $\lambda > 0$.

(3.3) Computing $\sigma(A)$. N. Linial, A. Samorodnitsky, and A. Wigderson present in [L+00] a deterministic polynomial time algorithm, which, given an $N \times N$ positive matrix A and a number $\epsilon > 0$ computes the value of $\sigma(A)$ within a factor of $(1+\epsilon)$ in time polynomial in $\ln \epsilon^{-1}$ and N (in the unit cost model).

We are interested in computing $\sigma(A)$ where $A = A(\gamma)$ is a random matrix of Theorem 1.2. Thus A is positive with probability 1. We observe that we can further save on computations as follows.

Let us consider the $m \times n$ matrix $(w_{ij}\gamma_{ij})$, which is also positive with probability 1. Applying the algorithm of [L+00], we can scale the matrix to the row sums r_i and the column sums c_j. Namely, we can compute (approximately, in polynomial time) positive numbers λ_i, $i = 1, \ldots, m$, and μ_j, $j = 1, \ldots, n$, and an $m \times n$ positive matrix $L = (l_{ij})$ such that

$$w_{ij}\gamma_{ij} = l_{ij}\mu_i\lambda_j \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n$$

and such that

$$\sum_{j=1}^{n} l_{ij} = r_i \quad \text{for} \quad i = 1, \ldots, m \quad \text{and}$$

$$\sum_{i=1}^{m} l_{ij} = c_j \quad \text{for} \quad j = 1, \ldots, n.$$

If we divide every row of A from R_i by $\mu_i r_i$ and divide every column from C_j by $\lambda_j c_j$, we get the $N \times N$ matrix with the entries in the $R_i \times C_j$ block equal to $l_{ij}/r_i c_j$. It is seen that the obtained matrix is doubly stochastic. Therefore, we have

$$\sigma(A) = \left(\prod_{i=1}^{m} (\mu_i r_i)^{r_i} \right) \left(\prod_{j=1}^{n} (\lambda_j c_j)^{c_j} \right).$$

Hence the scaling of the $N \times N$ matrix A reduces to the scaling of the $m \times n$ matrix $(w_{ij}\gamma_{ij})$.

11

4. Integrating $\sigma(A)$

Here we describe an algorithm for computing

$$T'(R, C; W) = \frac{N!}{N^N} \frac{\mathbf{E}\,\sigma(A)}{r_1! \cdots r_m! c_1! \cdots c_n!},$$

cf. Section 1.3.

(4.1) Notation. We interpret the space $\mathbb{R}^{mn}$ as the space of all $m \times n$ matrices $\gamma = (\gamma_{ij})$. Let $\mathbb{R}_+^{mn}$ denote the positive orthant $\gamma_{ij} > 0$ of $\mathbb{R}^{mn}$ and let

$$\Delta = \left\{ \gamma : \quad \sum_{ij} \gamma_{ij} = 1 \quad \text{and} \quad \gamma_{ij} > 0 \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n \right\}$$

be the standard (open) simplex in $\mathbb{R}^{mn}$.

For $0 < \delta < 1/mn$ let us consider the δ-*interior* of Δ:

$$\Delta_\delta = \left\{ \gamma : \quad \sum_{ij} \gamma_{ij} = 1 \quad \text{and} \quad \gamma_{ij} > \delta \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n \right\}.$$

Geometrically, Δ_δ is an open simplex lying strictly inside Δ.

For a $\tau \in \mathbb{R}$, let ν be the Lebesgue measure on the affine hyperplane

$$\sum_{ij} \gamma_{ij} = \tau$$

induced by the Euclidean structure on $\mathbb{R}^{mn}$.

For a matrix $\gamma \in \mathbb{R}_+^{mn}$, let

$$P(\gamma) = \operatorname{per} A(\gamma) \quad \text{and let} \quad S(\gamma) = \sigma\left(A(\gamma)\right),$$

where $A(\gamma)$ is the matrix constructed in Theorem 1.2 and σ is the function of Theorem 3.1.

Thus we have

$$T(R, C; W) = \frac{1}{r_1! \cdots r_m! c_1! \cdots c_n!} \int_{\mathbb{R}_+^{mn}} P(\gamma) \exp\left\{ -\sum_{ij} \gamma_{ij} \right\} d\gamma$$

and

$$T'(R, C; W) = \frac{N!}{N^N} \frac{1}{r_1! \cdots r_m! c_1! \cdots c_n!} \int_{\mathbb{R}_+^{mn}} S(\gamma) \exp\left\{ -\sum_{ij} \gamma_{ij} \right\} d\gamma,$$

where $d\gamma$ is the Lebesgue measure on $\mathbb{R}^{mn}$.

12

To apply the results of [AK91], [F+94], [FK99] (see also [Ve05]) on efficient integration of log-concave functions, we modify the problem to that of integration of $P(\gamma)$ and $S(\gamma)$ first on Δ and then on Δ_δ.

We use that both functions $P(\gamma)$ and $S(\gamma)$ are positive homogeneous of degree N and monotone on $\mathbb{R}_+^{mn}$: if $\gamma = (\gamma_{ij})$ and $\gamma' = (\gamma'_{ij})$ are positive matrices such that

$$\gamma'_{ij} \leq \gamma_{ij} \quad \text{for all} \quad i, j,$$

then

$$P(\gamma') \leq P(\gamma) \quad \text{and} \quad S(\gamma') \leq S(\gamma),$$

cf. Remark 3.2.

(4.2) Lemma. *We have*

$$\int_{\mathbb{R}_+^{mn}} P(\gamma) \exp\Big\{ -\sum_{ij} \gamma_{ij} \Big\} d\gamma = \frac{(N + mn - 1)!}{\sqrt{mn}} \int_{\Delta} P(\gamma) \, d\nu(\gamma).$$

Proof. We note that

$$\mathbb{R}_+^{mn} = \bigcup_{\tau > 0} \tau\Delta.$$

Since $d\nu \, d\tau = \sqrt{mn} \, d\gamma$, we get

$$\int_{\mathbb{R}_+^{mn}} P(\gamma) \exp\Big\{ -\sum_{ij} \gamma_{ij} \Big\} d\gamma = \frac{1}{\sqrt{mn}} \int_0^{+\infty} e^{-\tau} \left(\int_{\tau\Delta} P(\gamma) \, d\nu(\gamma) \right) d\tau.$$

Since $P(\gamma)$ is positive homogeneous of degree N, we conclude that

$$\int_{\tau\Delta} P(\gamma) \, d\nu(\gamma) = \tau^{N+mn-1} \int_{\Delta} P(\gamma) \, d\nu(\gamma),$$

from which the proof follows. $\square$

The same identity holds for the integrals of $S(\gamma)$.

Next, we approximate the integral over the simplex Δ by the integral over the inner simplex Δ_δ.

(4.3) Lemma. *Let $\delta \leq 1/mn$ be a non-negative number. Then*

$$\left(1 - mn\delta \right)^{N+mn-1} \int_{\Delta} P(\gamma) \, d\nu(\gamma) \leq \int_{\Delta_\delta} P(\gamma) \, d\nu(\gamma) \leq \int_{\Delta} P(\gamma) \, d\nu(\gamma).$$

Proof. To prove the lower bound, we observe that the transformation

$$\gamma_{ij} \longmapsto \gamma_{ij} - \delta \quad \text{for} \quad i = 1, \dots, m \quad \text{and} \quad j = 1, \dots, n$$

13

maps Δ_δ inside $(1 - \delta mn)\Delta$. Since P is monotone, we get

$$\int_{\Delta_\delta} P(\gamma)\, d\nu(\gamma) \geq \int_{(1-\delta mn)\Delta} P(\gamma)\, d\nu(\gamma) = (1 - \delta mn)^{N+mn-1} \int_\Delta P(\gamma)\, d\nu(\gamma),$$

where we used that P is homogeneous of degree N.

The upper bound is obvious. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

The same inequalities hold for the integrals of $S(\gamma)$.

For an $0 < \epsilon < 1$, let us choose a positive

$$\delta \leq \frac{-\ln(1 - \epsilon)}{mn(N + mn - 1)} \approx \frac{\epsilon}{mn(N + mn - 1)} \quad \text{for small} \quad \epsilon > 0.$$

Then the integral

$$\int_{\Delta_\delta} P(\gamma)\, d\nu(\gamma)$$

approximates the integral

$$\int_\Delta P(\gamma)\, d\nu(\gamma)$$

within a factor of $(1 - \epsilon)$ and the same holds for the integrals of $S(\gamma)$.

Since $A(\gamma)$ depends linearly on γ, by the results of Section 3, $S(\gamma)$ is a strictly positive log-concave function on the set of positive matrices γ and the value of $S(\gamma)$ can be computed in polynomial time for any given positive matrix γ.

Our goal consists of estimating $T(R, C; W)$ by

$$T'_\delta(R, C; W) = \frac{N!}{N^N} \frac{1}{r_1! \cdots r_m! c_1! \cdots c_n!} \frac{(N + mn - 1)!}{\sqrt{mn}} \int_{\Delta_\delta} S(\gamma)\, d\gamma.$$

To compute the integral, we apply the algorithms of [AK91], [F+94], and [FK99]. The computational complexity of the algorithms is polynomial in the dimension $mn - 1$ of the integral *and* the Lipschitz constant of $\ln S$ on Δ_δ. Hence it remains to estimate the Lipschitz constant of $\ln S$.

(4.4) Lemma. *Let $\delta < 1/mn$ be a positive number. Let $\gamma = (\gamma_{ij})$ and $\gamma' = (\gamma'_{ij})$ be two matrices such that*

$$\gamma_{ij}, \gamma'_{ij} \geq \delta \quad \text{for all} \quad i, j.$$

Then

$$|\ln S(\gamma) - \ln S(\gamma')| \leq \frac{N}{\delta} \max_{ij} |\gamma_{ij} - \gamma'_{ij}|.$$

Proof. For $t = (\tau_1, \ldots, \tau_N)$, let

$$f_\gamma(t) = \sum_{p=1}^N \ln \left(\sum_{q=1}^N a_{pq}(\gamma) e^{\tau_q} \right), \quad \text{where} \quad A(\gamma) = (a_{pq}(\gamma))$$

14

is the matrix of Theorem 1.2. Letting

$$H = \left\{ (\tau_1, \dots, \tau_N) : \quad \sum_{i=1}^{N} \tau_i = 0 \right\},$$

by formula (3.1.4), we can write

(4.4.1)
$$\ln S(\gamma) = \min_{t \in H} f_\gamma(t).$$

Let
$$\alpha = \max_{ij} |\gamma_{ij} - \gamma'_{ij}|.$$

Then
$$\gamma_{ij} \le \gamma'_{ij} + \alpha \le \gamma'_{ij} \left(1 + \frac{\alpha}{\delta} \right) \quad \text{for all} \quad i, j$$

and, similarly,
$$\gamma'_{ij} \le \gamma_{ij} + \alpha \le \gamma_{ij} \left(1 + \frac{\alpha}{\delta} \right) \quad \text{for all} \quad i, j.$$

Since
$$a_{pq}(\gamma) = w_{ij}\gamma_{ij} \quad \text{provided} \quad p \in R_i \quad \text{and} \quad q \in C_j,$$

we have
$$a_{pq}(\gamma) \le a_{pq}(\gamma') \left(1 + \frac{\alpha}{\delta} \right)$$

and, similarly,
$$a_{pq}(\gamma') \le a_{pq}(\gamma) \left(1 + \frac{\alpha}{\delta} \right).$$

Therefore, for all $t = (\tau_1, \dots, \tau_N)$, we have

$$f_\gamma(t) \le f_{\gamma'}(t) + N \ln \left(1 + \frac{\alpha}{\delta} \right) \le f_{\gamma'}(t) + \frac{\alpha N}{\delta}$$

and, similarly,
$$f_{\gamma'}(t) \le f_\gamma(t) + \frac{\alpha N}{\delta}.$$

Applying (4.4.1), we complete the proof. $\qquad\qquad\square$

Summarizing, we conclude that there is a randomized algorithm, which, for any given $\epsilon > 0$ computes the value of

$$T'(R, C; W) = \frac{N!}{N^N} \frac{\mathbf{E}\,\sigma(A)}{r_1! \cdots r_m! c_1! \cdots c_n!},$$

where A is the matrix of Theorem 1.2, within relative error ϵ in time polynomial in ϵ^{-1} and N (in the unit cost model).

15

5. Proof of Theorem 1.4

Our proof is based on two estimates for the permanent of a non-negative matrix.

(5.1) The van der Waerden bound. Let $B = (b_{ij})$ be an $N \times N$ doubly stochastic matrix, that is, a non-negative matrix with all row and column sums equal to 1. Then

$$\operatorname{per} B \geq \frac{N!}{N^N}.$$

This bound constituted van B.L. der Waerden's conjecture proved by G.P. Egorychev [Eg81] and D.I. Falikman [Fa81], see also Chapter 12 of [LW01].

(5.2) A continuous extension of the Minc-Bregman bound. Let $B = (b_{ij})$ be an $N \times N$ non-negative matrix. Let

$$s_i = \sum_{j=1}^{N} b_{ij} \quad \text{for} \quad i = 1, \dots, N$$

be the row sums of B.

If $b_{ij} \in \{0, 1\}$, the bound

$$\operatorname{per} B \leq \prod_{i=1}^{N} (s_i!)^{1/s_i}$$

was conjectured by H. Minc and proved by L.M. Bregman [Br73], see also Chapter 2 of [AS00].

A. Samorodnitsky communicated to the author the following extension of the Minc-Bregman bound. Suppose that

$$(5.2.1) \qquad \sum_{j=1}^{N} b_{ij} = 1 \quad \text{for all} \quad \text{for} \quad i = 1, \dots, N$$

and that

$$(5.2.2) \qquad b_{ij} \leq 1/t_i \quad \text{for all} \quad i = 1, \dots, N,$$

where $t_i, i = 1, \dots, N$, are positive integers. Then

$$(5.2.3) \qquad \operatorname{per} B \leq \prod_{i=1}^{N} \frac{(t_i!)^{1/t_i}}{t_i}.$$

To deduce (5.2.3), we argue that the maximum of $\operatorname{per} B$ on the class of $N \times N$ non-negative matrices satisfying (5.2.1) and (5.2.2) is attained at a matrix with $b_{ij} \in \{0, 1/t_i\}$ for all i, j. Indeed, let us choose a particular row index i. Then

16

any non-negative matrix B satisfying (5.2.1)–(5.2.2) can be written as a convex combination of two non-negative matrices B' and B'' which satisfy (5.2.1)–(5.2.2), agree with B in all rows, except possibly the ith row, and, additionally, satisfy $b'_{ij}, b''_{ij} \in \{0, 1/t_i\}$. Since the function per B is linear in every row, we conclude that per $B \leq \max\{\mathrm{per}(B'), \mathrm{per}(B'')\}$. Proceeding as above for rows $i = 1, \ldots, N$, we may assume that $b_{ij} \in \{0, 1/t_i\}$ for all i, j, so (5.2.3) follows from the Minc-Bregman bound.

A similar bound is obtained by G.W. Soules [So03]. If B is a non-negative matrix satisfying (5.2.1)–(5.2.2) where t_i do not have to be integer, then

$$\mathrm{per}\, B \leq \prod_{i=1}^{N} \frac{\Gamma^{1/t_i}(t_i + 1)}{t_i}.$$

Proof of Theorem 1.4. For a given $m \times n$ positive matrix $\gamma = (\gamma_{ij})$, let $A = A(\gamma)$ be the matrix constructed in Theorem 1.2 and let $B = B(A)$ be the matrix constructed in Theorem 3.1. We have

$$\mathrm{per}\, A = \sigma(A)\, \mathrm{per}\, B,$$

and we estimate per B.

By the estimate of Section 5.1, we have

$$\mathrm{per}\, A \geq \frac{N!}{N^N} \sigma(A),$$

from which

$$T(R, C; W) \geq T'(R, C; W).$$

On the other hand, as is discussed in Section 3.3, we can construct $B = B(\gamma)$ as follows: first, we construct a positive $m \times n$ matrix $L = (l_{ij})$ such that

$$w_{ij}\gamma_{ij} = l_{ij}\mu_i\lambda_j \quad \text{for all} \quad i, j$$

and some positive numbers $\mu_1, \ldots, \mu_m$ and $\lambda_1, \ldots, n$ and such that

$$\sum_{j=1}^{n} l_{ij} = r_i \quad \text{for} \quad i = 1, \ldots, m$$

and

$$\sum_{i=1}^{m} l_{ij} = c_j \quad \text{for} \quad j = 1, \ldots, n$$

and then fill the $R_i \times C_j$ block of B by $l_{ij}/r_i c_j$.

It follows then that every entry of B in the block R_i of rows does not exceed $1/r_i$. Applying the bound of Section 5.2, we get

$$\mathrm{per}\, B \leq \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}}.$$

17

Similarly, every entry of B in the block C_j of columns does not exceed $1/c_j$, so we get

$$\text{per } B \leq \prod_{i=1}^{n} \frac{c_j!}{c_j^{c_j}}.$$

Therefore,

$$\text{per } A \leq \sigma(A) \min \left\{ \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}}, \quad \prod_{i=1}^{n} \frac{c_j!}{c_j^{c_j}} \right\}.$$

and the proof follows. □

Acknowledgments

References

[AK91]   D. Applegate and R. Kannan, *Sampling and integration of log-concave functions*, Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing, ACM, 1991, pp. 156–163.

[AS00]   N. Alon and J. Spencer, *The Probabilistic Method. Second edition*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[Ba05]   A. Barvinok, *Low rank approximations of symmetric polynomials and asymptotic counting of contingency tables*, preprint arXiv math.CO/0503170 (2005).

[BP99]   A. Barvinok and J.E. Pommersheim, *An algorithmic theory of lattice points in polyhedra*, New Perspectives in Algebraic Combinatorics (Berkeley, CA, 1996–97), Math. Sci. Res. Inst. Publ., vol. 38, Cambridge Univ. Press, Cambridge, 1999, pp. 91–147.

[Br73]   L.M. Bregman, *Certain properties of nonnegative matrices and their permanents*, Dokl. Akad. Nauk SSSR **211** (1973), 27–30.

[B+04]   W. Baldoni-Silva, J.A. De Loera, and M. Vergne, *Counting integer flows in networks*, Found. Comput. Math. **4** (2004), 277–314.

[C+05]   M. Cryan, M. Dyer, and D. Randall, *Approximately counting integral flows and cell-bounded contingency tables*, STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing, ACM, New York, 2005, pp. 413–422.

[CD03]   M. Cryan and M. Dyer, *A polynomial-time algorithm to approximately count contingency tables when the number of rows is constant*, Special issue on STOC2002 (Montreal, QC), J. Comput. System Sci. **67** (2003), 291–310.

[DG95]   P. Diaconis and A. Gangolli, *Rectangular arrays with fixed margins*, Discrete Probability and Algorithms (Minneapolis, MN, 1993), IMA Vol. Math. Appl., vol. 72, Springer, New York, 1995, pp. 15–41.

[DG04]   P. Diaconis and A. Gamburd, *Random matrices, magic squares and matching polynomials*, Research Paper 2, pp. 26, Electron. J. Combin. **11** (2004/05).

[D+97]   M. Dyer, R. Kannan, and J. Mount, *Sampling contingency tables*, Random Structures Algorithms **10** (1997), 487–506.

[Eg81]   G.P. Egorychev, *The solution of van der Waerden's problem for permanents*, Adv. in Math. **42** (1981), 299–305.

[Fa81]   D.I. Falikman, *Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix (Russian)*, Mat. Zametki **29** (1981), 931–938.

[FK99]   A. Frieze and R. Kannan, *Log-Sobolev inequalities and sampling from log-concave distributions*, Ann. Appl. Probab. **9** (1999), 14–26.

[F+94]   A. Frieze, R. Kannan, and N. Polson, *Sampling from log-concave distributions*, Ann. Appl. Probab. **4** (1994), 812–837; correction, p. 1255.

[GS02]   L. Gurvits and A. Samorodnitsky, *A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary*, Discrete Comput. Geom. **27** (2002), 531–550.

[J+04]   M. Jerrum, A. Sinclair, and E. Vigoda, *A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries*, J. ACM **51** (2004), 671–697.

[KK96]   B. Kalantari and L. Khachiyan, *On the complexity of nonnegative-matrix scaling*, Linear Algebra Appl. **240** (1996), 87–103.

[Lo71]   D. London, *On matrices with a doubly stochastic pattern*, J. Math. Anal. Appl. **34** (1971), 648–652.

[LW01]   J.H. van Lint and R.M. Wilson, *A Course in Combinatorics. Second edition*, Cambridge University Press, Cambridge, 2001.

[L+00]   N. Linial, A. Samorodnitsky, and A. Wigderson, *A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents*, Combinatorica **20** (2000), 545–568.

[Mi78]   H. Minc, *Permanents*, Encyclopedia of Mathematics and its Applications, vol. 6, Addison-Wesley Publishing Co., Reading, Mass, 1978.

[Mo02]   B.J. Morris, *Improved bounds for sampling contingency tables*, Random Structures Algorithms **21** (2002), 135–146.

[NR99]   A. Nemirovski and U. Rothblum, *On complexity of matrix scaling*, Linear Algebra Appl. **302/303** (1999), 435–460.

[Sa06]   A. Samorodnitsky, personal communication (2006).

[Si64]   R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist. **35** (1964), 876–879.

[So03]   G.W. Soules, *New permanental upper bounds for nonnegative matrices*, Linear Multilinear Algebra **51** (2003), 319–337.

[St97]   R.P. Stanley, *Enumerative Combinatorics. Vol. 1. Corrected reprint of the 1986 original*, Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, Cambridge, 1997.

[Ve05]   S. Vempala, *Geometric random walks: a survey*, Combinatorial and Computational Geometry, Math. Sci. Res. Inst. Publ., vol. 52, Cambridge Univ. Press, Cambridge, 2005, pp. 577–616.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043
E-mail address: `barvinok@umich.edu`